

Assignment #6

CSCI 581, Spring 2022

Jayaa Emekar

Identifying handwritten digits

The [Digits](#) dataset in `sklearn` has 1,797 samples with 64 numerical features (8×8 pixels) and a 10-class target variable representing the digits 0 through 9.

All dependencies

This is all the imports needed for this notebook

```
In [ ]: import pandas as pd
import seaborn as sns; sns.set()
import matplotlib.pyplot as plt
import time
from sklearn.datasets import load_digits
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import classification_report, multilabel_confusion_matrix
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression, LogisticRegressionCV
from sklearn.metrics import confusion_matrix
import seaborn as sns; sns.set()

import warnings
warnings.filterwarnings("ignore")
```

Explore dataset

to get as better understanding of the dataset's content

```
In [ ]: # Load the Digits dataset
digits = load_digits(as_frame=1)

X = digits.data
y = digits.target

# overview of data
print('\n', digits.data.info())

# Check shape of X
print('\nX.shape =', X.shape)

# Each image is 8px by 8px that's why 64 pixels
print('\ndigits.images[4] =')
print(digits.images[4])

# Let's check a few particular target values
targetValues = [22, 59, 32, 11]

for val in targetValues:
    print(f'\ny[{val}] = {y[val]}')

    # Next, let's see how the image looks like
    plt.gray()
    plt.matshow(digits.images[val])
    plt.show()
    plt.clf()

# check all unique possible target values
print(f'\nPossible target values = {set(digits.target)}')

# visualize any potential bias in the dataset towards a specific target value
sns.countplot(digits.target)
plt.show()
plt.clf()

# print count of each value
print(digits.target.value_counts())
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1797 entries, 0 to 1796
```

```
Data columns (total 64 columns):
```

#	Column	Non-Null Count	Dtype
0	pixel_0_0	1797 non-null	float64
1	pixel_0_1	1797 non-null	float64
2	pixel_0_2	1797 non-null	float64
3	pixel_0_3	1797 non-null	float64
4	pixel_0_4	1797 non-null	float64
5	pixel_0_5	1797 non-null	float64
6	pixel_0_6	1797 non-null	float64
7	pixel_0_7	1797 non-null	float64
8	pixel_1_0	1797 non-null	float64
9	pixel_1_1	1797 non-null	float64
10	pixel_1_2	1797 non-null	float64
11	pixel_1_3	1797 non-null	float64
12	pixel_1_4	1797 non-null	float64
13	pixel_1_5	1797 non-null	float64
14	pixel_1_6	1797 non-null	float64
15	pixel_1_7	1797 non-null	float64
16	pixel_2_0	1797 non-null	float64
17	pixel_2_1	1797 non-null	float64
18	pixel_2_2	1797 non-null	float64
19	pixel_2_3	1797 non-null	float64
20	pixel_2_4	1797 non-null	float64
21	pixel_2_5	1797 non-null	float64
22	pixel_2_6	1797 non-null	float64
23	pixel_2_7	1797 non-null	float64
24	pixel_3_0	1797 non-null	float64
25	pixel_3_1	1797 non-null	float64
26	pixel_3_2	1797 non-null	float64
27	pixel_3_3	1797 non-null	float64
28	pixel_3_4	1797 non-null	float64
29	pixel_3_5	1797 non-null	float64
30	pixel_3_6	1797 non-null	float64
31	pixel_3_7	1797 non-null	float64
32	pixel_4_0	1797 non-null	float64
33	pixel_4_1	1797 non-null	float64
34	pixel_4_2	1797 non-null	float64
35	pixel_4_3	1797 non-null	float64
36	pixel_4_4	1797 non-null	float64
37	pixel_4_5	1797 non-null	float64
38	pixel_4_6	1797 non-null	float64
39	pixel_4_7	1797 non-null	float64

```
40 pixel_5_0 1797 non-null float64
41 pixel_5_1 1797 non-null float64
42 pixel_5_2 1797 non-null float64
43 pixel_5_3 1797 non-null float64
44 pixel_5_4 1797 non-null float64
45 pixel_5_5 1797 non-null float64
46 pixel_5_6 1797 non-null float64
47 pixel_5_7 1797 non-null float64
48 pixel_6_0 1797 non-null float64
49 pixel_6_1 1797 non-null float64
50 pixel_6_2 1797 non-null float64
51 pixel_6_3 1797 non-null float64
52 pixel_6_4 1797 non-null float64
53 pixel_6_5 1797 non-null float64
54 pixel_6_6 1797 non-null float64
55 pixel_6_7 1797 non-null float64
56 pixel_7_0 1797 non-null float64
57 pixel_7_1 1797 non-null float64
58 pixel_7_2 1797 non-null float64
59 pixel_7_3 1797 non-null float64
60 pixel_7_4 1797 non-null float64
61 pixel_7_5 1797 non-null float64
62 pixel_7_6 1797 non-null float64
63 pixel_7_7 1797 non-null float64
```

```
dtypes: float64(64)
```

```
memory usage: 898.6 KB
```

```
None
```

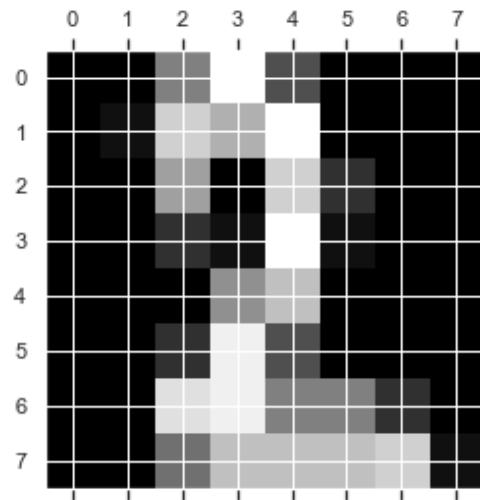
```
X.shape = (1797, 64)
```

```
digits.images[4] =
```

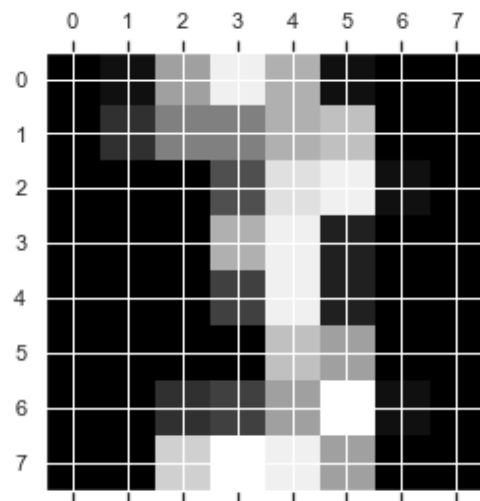
```
[[ 0.  0.  0.  1. 11.  0.  0.  0.]
 [ 0.  0.  0.  7.  8.  0.  0.  0.]
 [ 0.  0.  1. 13.  6.  2.  2.  0.]
 [ 0.  0.  7. 15.  0.  9.  8.  0.]
 [ 0.  5. 16. 10.  0. 16.  6.  0.]
 [ 0.  4. 15. 16. 13. 16.  1.  0.]
 [ 0.  0.  0.  3. 15. 10.  0.  0.]
 [ 0.  0.  0.  2. 16.  4.  0.  0.]]
```

```
y[22] = 2
```

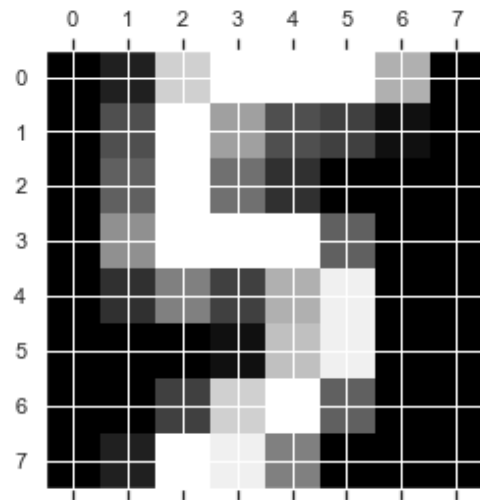
```
<Figure size 432x288 with 0 Axes>
```



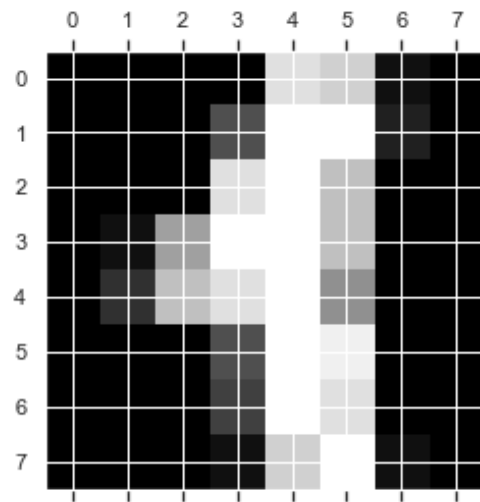
y[59] = 3
<Figure size 432x288 with 0 Axes>



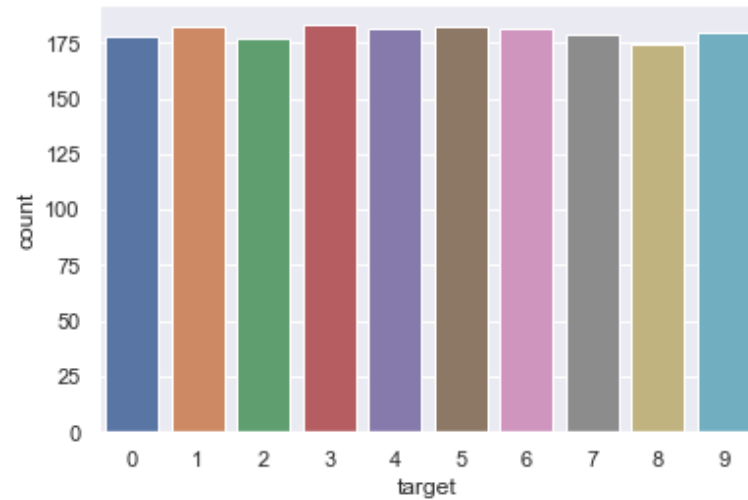
y[32] = 5
<Figure size 432x288 with 0 Axes>



y[11] = 1
 <Figure size 432x288 with 0 Axes>



Possible target values = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}



```

3    183
1    182
5    182
4    181
6    181
9    180
7    179
0    178
2    177
8    174

```

Name: target, dtype: int64

<Figure size 432x288 with 0 Axes>

Observation

- The data set includes a grey scale of pixels (8 by 8, or 46 pixels per image).
- each image represents a hand written integer. The integer is one of the following possible values:
 - Possible target values = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}
- Each column (pixel position, per image) includes 1797 non-null values of type `float64`.
- The data set appears to have no specific bias towards a specific integer. the integer `3` has the most number samples (183), where the integer `8` has the least number of samples (174).

Split the dataset for training and testing

The same test and training sub-datasets will be used for all models to ensure a normalized comparison.

```
In [ ]: # To apply an estimator (classifier) on this data, we flatten the image,
# to transform the data into a (samples, feature) matrix:
n_samples = len(digits.images)
X = digits.images.reshape((n_samples, -1))
y = digits.target

# Split the dataset in two equal parts
X_train, X_test, y_train, y_test = train_test_split(
    X ,          # (samples, features) matrix
    y ,          # target values
    test_size=0.5 , # 50-50 split
    stratify = y,  # stratify to maintain original distribution of possible target values
    random_state=23 # constant value to ensure repeatability
)
```

Helper functions

These function will be used to aid visualization throughout the nbotebook.

```
In [ ]: def addSubPlotConfusionMatrix(confMatrix, axes, target):
    ''' add a subplot in the form of a confusion matrix '''

    # conver to a df to easily plot it via seaborn as a heatmap
    df_cm = pd.DataFrame(confMatrix, index=['T', 'F'],
                          columns=['P', 'N'])

    # create sub-plot object to be added to the high-level plot object
    subplot = sns.heatmap(df_cm, annot=True, square=True,
                          cbar=False, fmt="d", ax=axes)

    # set up axis, title and labels of sub-plot at hand
    subplot.xaxis.set_ticklabels(subplot.xaxis.get_ticklabels(),
                                 ha='right', fontsize=16)
    subplot.yaxis.set_ticklabels(subplot.yaxis.get_ticklabels(),
                                 ha='right', fontsize=16)
    axes.set_title(f"Class {target}")
    axes.set_xlabel('Model Prediction')
    axes.set_ylabel('Ground Truth')
```



```
return
```

Part 1: Classification using a support vector machine

Generate a classification system for this problem using a *Support Vector Machine (SVM)*. You may refer to the `sklearn_grid_search_digits.ipynb` Jupyter notebook example we discussed in class for hyperparameter settings for this estimator.

Solution

first, let's create an SVM classifier without grid search and see how it performs with minimal variables specified. Here we are only specifying the kernel type as linear.

```
In [ ]: '''
first, let's create an SVM classifier without grid search and see how it performs
with minimal variables specified. Here we are only specifying the kernel type as linear.
'''

# create model with linear kernel
classifier = SVC(kernel='linear')

# train model
classifier.fit(X_train,y_train)

# predict target values using our trained model
y_pred = classifier.predict(X_test)

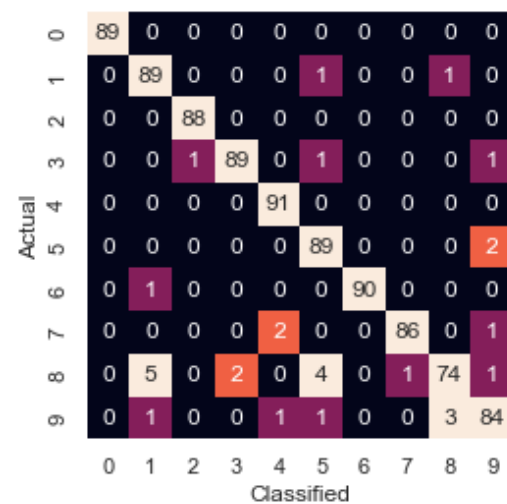
# show classification report
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.93	0.98	0.95	91
2	0.99	1.00	0.99	88
3	0.98	0.97	0.97	92
4	0.97	1.00	0.98	91
5	0.93	0.98	0.95	91
6	1.00	0.99	0.99	91
7	0.99	0.97	0.98	89
8	0.95	0.85	0.90	87
9	0.94	0.93	0.94	90
accuracy			0.97	899
macro avg	0.97	0.97	0.97	899
weighted avg	0.97	0.97	0.97	899

Plot Confusion Matrix

```
In [ ]: expected = y_test
predicted = classifier.predict(X_test)

mat = confusion_matrix(expected, predicted)
sns.heatmap(mat, square=True, annot=True, fmt='d', cbar=False, vmax=3)
plt.xlabel('Classified')
plt.ylabel('Actual');
```



Due to the high accuracy of this model, I set the vmax parameter on the heatmap below to 3 to make it easier to differentiate 1s from 0s. Consequently, any value above 3 will appear as the same color.

This is remarkably accurate. For a more useful visual, and because this runs fairly quickly, we see how a stack of these predictions with different random seeds would add up.

Now let's try to create the same type of classifier, but utilize grid search to tune the classifier's hyperparameters

Use grid search parameters range and search cases

See all possible parameters and their types, possible combinations, and their impact

```
In [ ]: '''
now let's try to create the same type of classifier, but utilize grid search to tune
the classifier's hyperparameters
'''

# grid search parameters range and search cases
# See all possible parameters and their types, possible combinations, and their impact:
# https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html
searchCases = [
    {"kernel": ["rbf"],
     "gamma": [1e-3, 1e-4], # default=0.0
     "C": [1, 10, 100, 1000], # default=1.0
     "shrinking": [True, False] # default=True
    },
    {"kernel": ["poly"],
     "gamma": [1e-3, 1e-4], # default=0.0
     "C": [1, 10, 100, 1000], # default=1.0
     "degree": [1, 2, 3], # default=3
     "coef0": [0.0, 0.1, 1.0, 10.0], # default=0.0
     "shrinking": [True, False] # default=True
    },
    {"kernel": ["linear"],
     "C": [1, 10, 100, 1000], # default=1.0
     "shrinking": [True, False] # default=True
    },
    {"kernel": ["sigmoid"],
     "C": [1, 10, 100, 1000], # default=1.0
     "degree": [1, 2, 3], # default=3
     "coef0": [0.0, 0.1, 1.0], # default=0.0
    }
```

```

        "shrinking": [True, False] # default=True
    }
]

```

These are the metrics we will use to rank each grid search result and find the best possible outcome

```

In [ ]: # these are the metrics we will use to rank each grid search result and
# find the best possible outcome
metrics = ["precision", "recall", "f1", "jaccard"]
allResultsAllMetrics = []

for metric in metrics:
    allResultsPerMetric = []
    print(f"Finding the best hyperparameters for our SVC classifier based on the best {metric} score")

    # create model, and use grid search to tune the hyperparameters
    classifier = GridSearchCV(SVC(), searchCases, scoring=f"{metric}_macro")

    # train model
    classifier.fit(X_train, y_train)

    # display best results found via grid search
    print(f"Best parameters set found on development set:\n{classifier.best_params_}\n\n")

    # display best parameters combination
    means = classifier.cv_results_["mean_test_score"]
    stds = classifier.cv_results_["std_test_score"]
    for mean, std, params in zip(means, stds, classifier.cv_results_["params"]):
        allResultsPerMetric.append("%0.3f (+/-%0.03f) for %r" % (mean, std * 2, params))

    # display detailed report of best case
    print("Detailed classification report:\n")
    print("The model is trained on the full development set.\n",
          "The scores are computed on the full evaluation set.\n")
    y_true, y_pred = y_test, classifier.predict(X_test)
    print(classification_report(y_true, y_pred), '\n\n')

    # plot confusion matrix
    fig, ax = plt.subplots(2, 5, figsize=(10, 6))
    for cfs_matrix, target, axes in zip(multilabel_confusion_matrix(y_true, y_pred),
                                        range(0,10),
                                        ax.flatten()):
        addSubPlotConfusionMatrix(cfs_matrix, axes, target)
    fig.suptitle(f'Multiclass Confusion Matrix for metric {metric}', fontsize=18)

```

```

fig.tight_layout(rect=[0, 0.03, 1, 0.95])
plt.show()
plt.clf()
print()

# append all result to the array to print it at the end
allResultsAllMetrics.append(allResultsPerMetric)

```

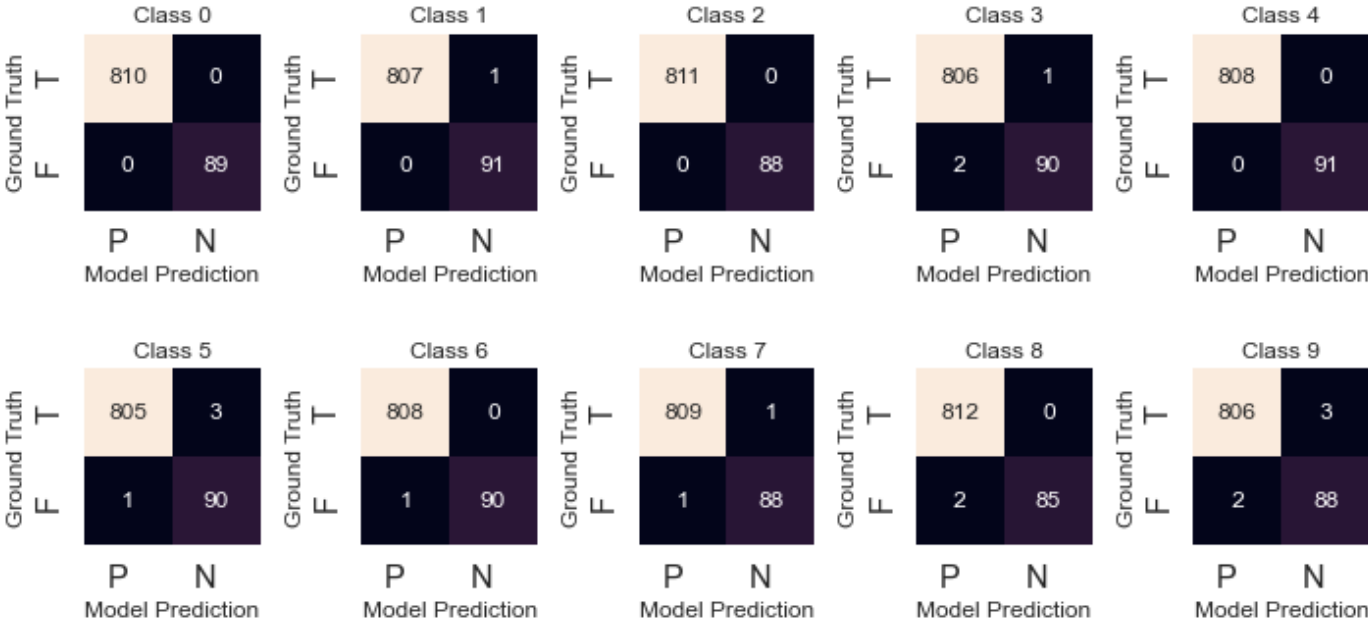
Finding the best hyperparameters for our SVC classifier based on the best precision score
 Best parameters set found on development set:
 {'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}

Detailed classification report:

The model is trained on the full development set.
 The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.99	1.00	0.99	91
2	1.00	1.00	1.00	88
3	0.99	0.98	0.98	92
4	1.00	1.00	1.00	91
5	0.97	0.99	0.98	91
6	1.00	0.99	0.99	91
7	0.99	0.99	0.99	89
8	1.00	0.98	0.99	87
9	0.97	0.98	0.97	90
accuracy			0.99	899
macro avg	0.99	0.99	0.99	899
weighted avg	0.99	0.99	0.99	899

Multiclass Confusion Matrix for metric precision



Finding the best hyperparameters for our SVC classifier based on the best recall score
Best parameters set found on development set:
{'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}

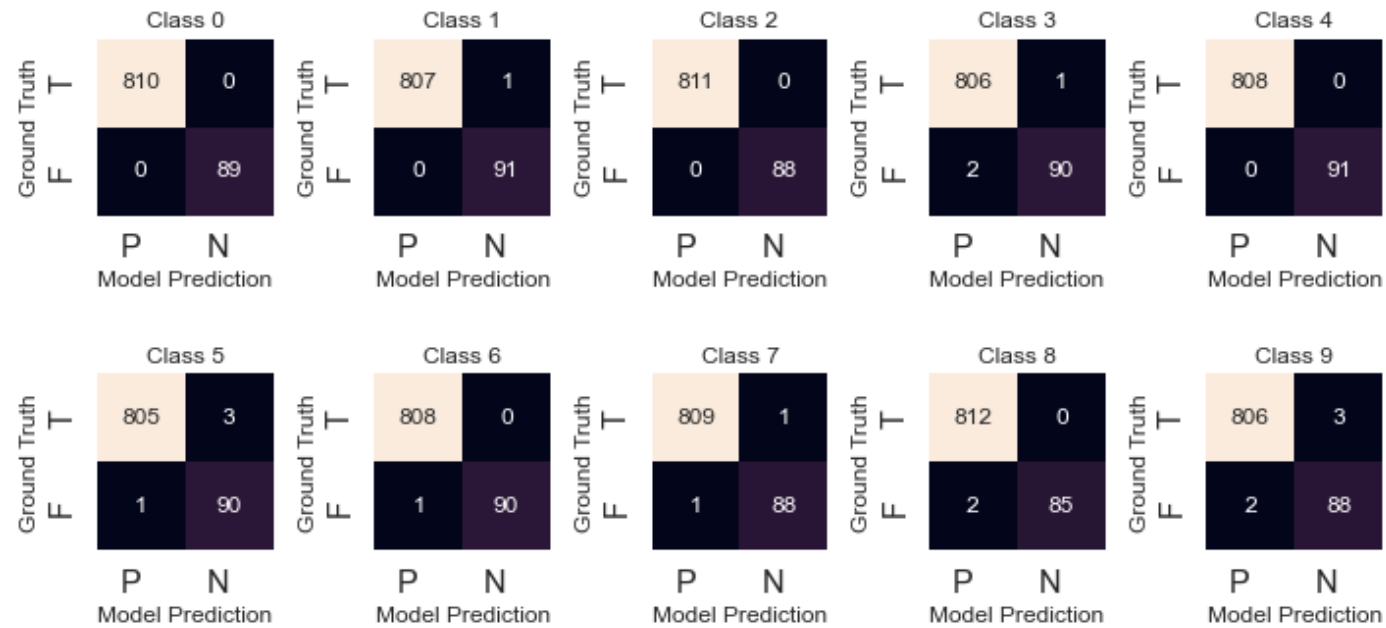
Detailed classification report:

The model is trained on the full development set.
The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.99	1.00	0.99	91
2	1.00	1.00	1.00	88
3	0.99	0.98	0.98	92
4	1.00	1.00	1.00	91
5	0.97	0.99	0.98	91
6	1.00	0.99	0.99	91
7	0.99	0.99	0.99	89
8	1.00	0.98	0.99	87
9	0.97	0.98	0.97	90
accuracy			0.99	899
macro avg	0.99	0.99	0.99	899
weighted avg	0.99	0.99	0.99	899

<Figure size 432x288 with 0 Axes>

Multiclass Confusion Matrix for metric recall



Finding the best hyperparameters for our SVC classifier based on the best f1 score
Best parameters set found on development set:
{'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}

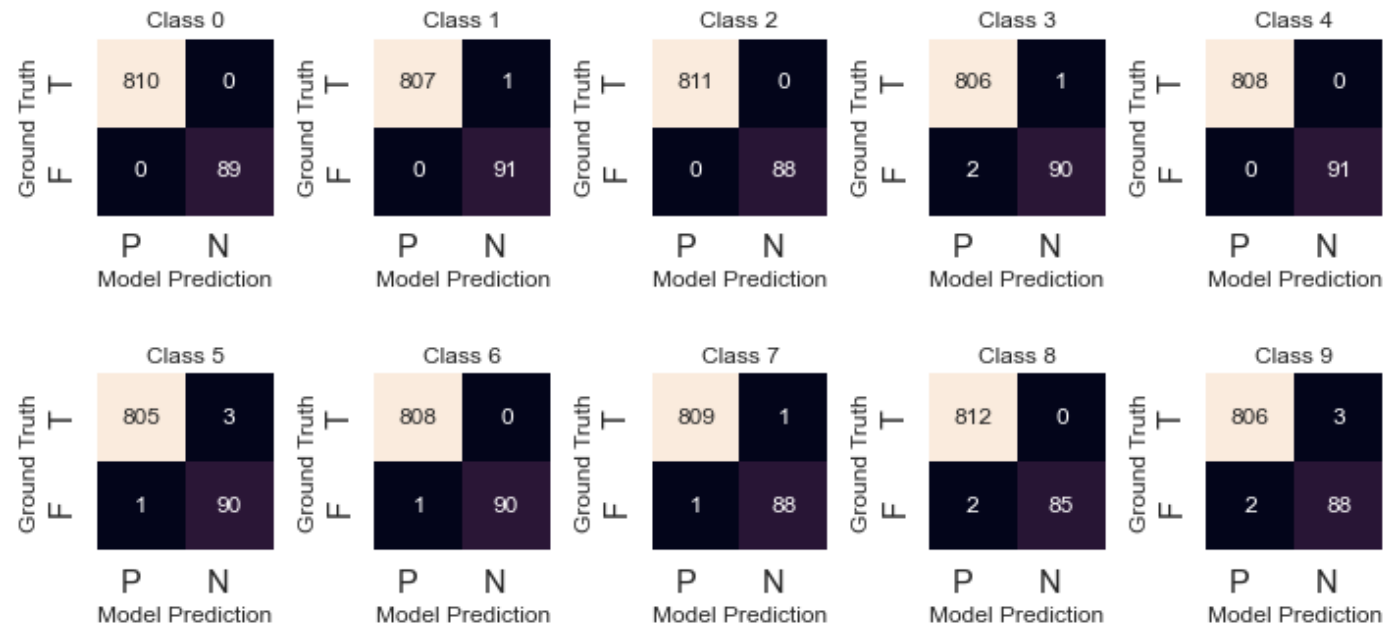
Detailed classification report:

The model is trained on the full development set.
The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.99	1.00	0.99	91
2	1.00	1.00	1.00	88
3	0.99	0.98	0.98	92
4	1.00	1.00	1.00	91
5	0.97	0.99	0.98	91
6	1.00	0.99	0.99	91
7	0.99	0.99	0.99	89
8	1.00	0.98	0.99	87
9	0.97	0.98	0.97	90
accuracy			0.99	899
macro avg	0.99	0.99	0.99	899
weighted avg	0.99	0.99	0.99	899

<Figure size 432x288 with 0 Axes>

Multiclass Confusion Matrix for metric f1



Finding the best hyperparameters for our SVC classifier based on the best jaccard score
Best parameters set found on development set:
{'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}

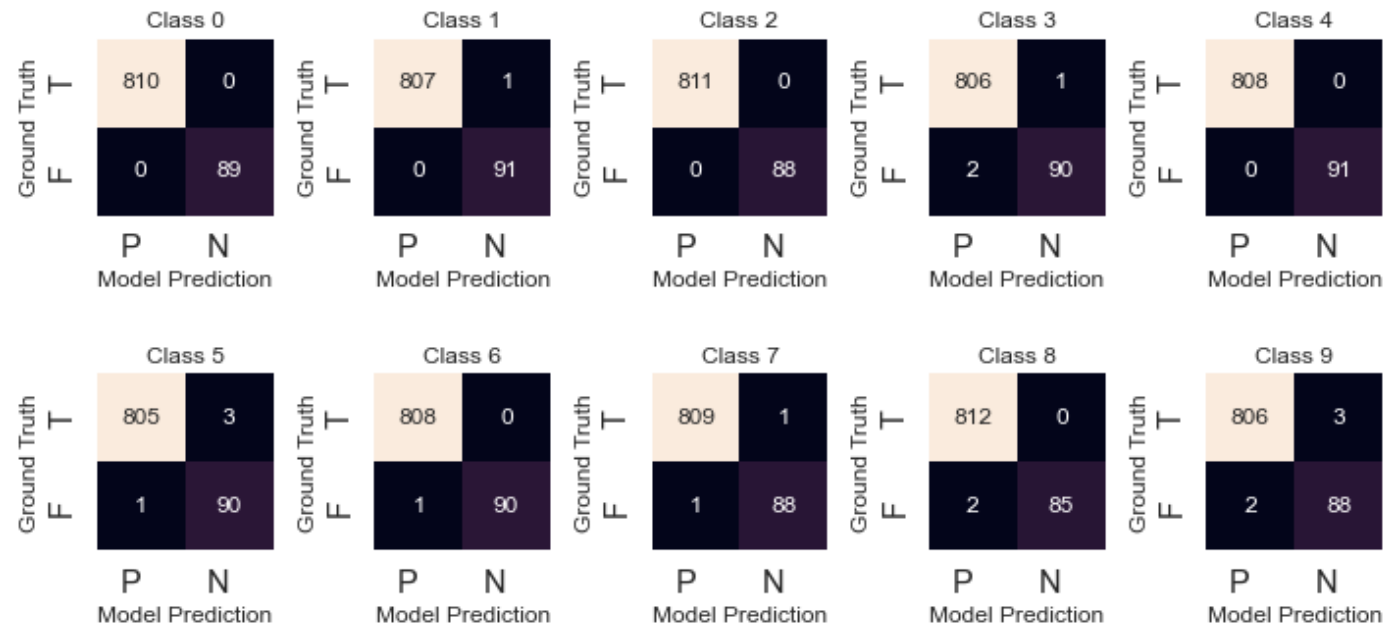
Detailed classification report:

The model is trained on the full development set.
The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.99	1.00	0.99	91
2	1.00	1.00	1.00	88
3	0.99	0.98	0.98	92
4	1.00	1.00	1.00	91
5	0.97	0.99	0.98	91
6	1.00	0.99	0.99	91
7	0.99	0.99	0.99	89
8	1.00	0.98	0.99	87
9	0.97	0.98	0.97	90
accuracy			0.99	899
macro avg	0.99	0.99	0.99	899
weighted avg	0.99	0.99	0.99	899

<Figure size 432x288 with 0 Axes>

Multiclass Confusion Matrix for metric jaccard



<Figure size 432x288 with 0 Axes>

Printed all results for all possible results from our grid search for SVC classifier based on precision, recall, f1 score and jaccard

```
In [ ]: # print all results (full verbosity)
for metric in range(len(allResultsAllMetrics)):
    print(f"All possible results from our grid search for SVC classifier based on {metrics[metric]}")
    for x in allResultsAllMetrics[metric]:
        print('\t', x)
    print()
```

All possible results from our grid search for SVC classifier based on precision

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

0.828 (+/-0.045) for {'C': 100, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.821 (+/-0.066) for {'C': 1000, 'coef0': 0.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.821 (+/-0.066) for {'C': 1000, 'coef0': 0.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.821 (+/-0.066) for {'C': 1000, 'coef0': 0.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.821 (+/-0.066) for {'C': 1000, 'coef0': 0.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.821 (+/-0.066) for {'C': 1000, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.821 (+/-0.066) for {'C': 1000, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.817 (+/-0.057) for {'C': 1000, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.817 (+/-0.057) for {'C': 1000, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.817 (+/-0.057) for {'C': 1000, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.817 (+/-0.057) for {'C': 1000, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.817 (+/-0.057) for {'C': 1000, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.817 (+/-0.057) for {'C': 1000, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.800 (+/-0.045) for {'C': 1000, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.800 (+/-0.045) for {'C': 1000, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.800 (+/-0.045) for {'C': 1000, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.800 (+/-0.045) for {'C': 1000, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.800 (+/-0.045) for {'C': 1000, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.800 (+/-0.045) for {'C': 1000, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}

```

All possible results from our grid search for SVC classifier based on recall

```

0.985 (+/-0.013) for {'C': 1, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.985 (+/-0.013) for {'C': 1, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.964 (+/-0.016) for {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.964 (+/-0.016) for {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.988 (+/-0.013) for {'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.988 (+/-0.013) for {'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.987 (+/-0.017) for {'C': 10, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.987 (+/-0.017) for {'C': 10, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.988 (+/-0.013) for {'C': 100, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.988 (+/-0.013) for {'C': 100, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.982 (+/-0.016) for {'C': 100, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.982 (+/-0.016) for {'C': 100, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.988 (+/-0.013) for {'C': 1000, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.988 (+/-0.013) for {'C': 1000, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.982 (+/-0.016) for {'C': 1000, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.982 (+/-0.016) for {'C': 1000, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.979 (+/-0.015) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': True}
0.979 (+/-0.015) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': False}
0.943 (+/-0.029) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': True}
0.943 (+/-0.029) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': False}
0.984 (+/-0.021) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': True}
0.984 (+/-0.021) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': False}
0.933 (+/-0.038) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': True}
0.933 (+/-0.038) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': False}

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

All possible results from our grid search for SVC classifier based on f1

```
0.985 (+/-0.014) for {'C': 1, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.985 (+/-0.014) for {'C': 1, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.964 (+/-0.016) for {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.964 (+/-0.016) for {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

0.789 (+/-0.069) for {'C': 1000, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.789 (+/-0.069) for {'C': 1000, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.784 (+/-0.059) for {'C': 1000, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.784 (+/-0.059) for {'C': 1000, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.784 (+/-0.059) for {'C': 1000, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.784 (+/-0.059) for {'C': 1000, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.784 (+/-0.059) for {'C': 1000, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.784 (+/-0.059) for {'C': 1000, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.760 (+/-0.048) for {'C': 1000, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.760 (+/-0.048) for {'C': 1000, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.760 (+/-0.048) for {'C': 1000, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.760 (+/-0.048) for {'C': 1000, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.760 (+/-0.048) for {'C': 1000, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.760 (+/-0.048) for {'C': 1000, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}

```

All possible results from our grid search for SVC classifier based on jaccard

```

0.972 (+/-0.026) for {'C': 1, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.972 (+/-0.026) for {'C': 1, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.933 (+/-0.028) for {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.933 (+/-0.028) for {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.976 (+/-0.025) for {'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.976 (+/-0.025) for {'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.974 (+/-0.031) for {'C': 10, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.974 (+/-0.031) for {'C': 10, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.976 (+/-0.025) for {'C': 100, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.976 (+/-0.025) for {'C': 100, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.966 (+/-0.030) for {'C': 100, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.966 (+/-0.030) for {'C': 100, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.976 (+/-0.025) for {'C': 1000, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}
0.976 (+/-0.025) for {'C': 1000, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}
0.966 (+/-0.030) for {'C': 1000, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': True}
0.966 (+/-0.030) for {'C': 1000, 'gamma': 0.0001, 'kernel': 'rbf', 'shrinking': False}
0.960 (+/-0.027) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': True}
0.960 (+/-0.027) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': False}
0.897 (+/-0.048) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': True}
0.897 (+/-0.048) for {'C': 1, 'coef0': 0.0, 'degree': 1, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': False}
0.970 (+/-0.040) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': True}
0.970 (+/-0.040) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': False}
0.879 (+/-0.065) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': True}
0.879 (+/-0.065) for {'C': 1, 'coef0': 0.0, 'degree': 2, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': False}
0.972 (+/-0.025) for {'C': 1, 'coef0': 0.0, 'degree': 3, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': True}
0.972 (+/-0.025) for {'C': 1, 'coef0': 0.0, 'degree': 3, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': False}
0.803 (+/-0.081) for {'C': 1, 'coef0': 0.0, 'degree': 3, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': True}
0.803 (+/-0.081) for {'C': 1, 'coef0': 0.0, 'degree': 3, 'gamma': 0.0001, 'kernel': 'poly', 'shrinking': False}
0.960 (+/-0.027) for {'C': 1, 'coef0': 0.1, 'degree': 1, 'gamma': 0.001, 'kernel': 'poly', 'shrinking': True}

```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

0.688 (+/-0.075) for {'C': 100, 'coef0': 0.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.688 (+/-0.075) for {'C': 100, 'coef0': 0.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.688 (+/-0.075) for {'C': 100, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.688 (+/-0.075) for {'C': 100, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.669 (+/-0.056) for {'C': 100, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.669 (+/-0.056) for {'C': 100, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.669 (+/-0.056) for {'C': 100, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.669 (+/-0.056) for {'C': 100, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.669 (+/-0.056) for {'C': 100, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.669 (+/-0.056) for {'C': 100, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.692 (+/-0.064) for {'C': 100, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.692 (+/-0.064) for {'C': 100, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.692 (+/-0.064) for {'C': 100, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.692 (+/-0.064) for {'C': 100, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.692 (+/-0.064) for {'C': 100, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.692 (+/-0.064) for {'C': 100, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.671 (+/-0.085) for {'C': 1000, 'coef0': 0.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.671 (+/-0.085) for {'C': 1000, 'coef0': 0.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.671 (+/-0.085) for {'C': 1000, 'coef0': 0.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.671 (+/-0.085) for {'C': 1000, 'coef0': 0.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.671 (+/-0.085) for {'C': 1000, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.671 (+/-0.085) for {'C': 1000, 'coef0': 0.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.663 (+/-0.075) for {'C': 1000, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.663 (+/-0.075) for {'C': 1000, 'coef0': 0.1, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.663 (+/-0.075) for {'C': 1000, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.663 (+/-0.075) for {'C': 1000, 'coef0': 0.1, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.663 (+/-0.075) for {'C': 1000, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.663 (+/-0.075) for {'C': 1000, 'coef0': 0.1, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}
0.632 (+/-0.064) for {'C': 1000, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': True}
0.632 (+/-0.064) for {'C': 1000, 'coef0': 1.0, 'degree': 1, 'kernel': 'sigmoid', 'shrinking': False}
0.632 (+/-0.064) for {'C': 1000, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': True}
0.632 (+/-0.064) for {'C': 1000, 'coef0': 1.0, 'degree': 2, 'kernel': 'sigmoid', 'shrinking': False}
0.632 (+/-0.064) for {'C': 1000, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': True}
0.632 (+/-0.064) for {'C': 1000, 'coef0': 1.0, 'degree': 3, 'kernel': 'sigmoid', 'shrinking': False}

```

Observation

We ran an extensive grid search where possible for our dataset. The grid search focused on tuning the different parameters for different kernel types of our SVM classifier. All other parameters seemed to influence the performance of the SVM model based on the kernel type.

For example, the parameter `gamma` has the most impact on SVM models with `rbf` or `poly` kernels, and the parameter `degree` has the most impact on SVM models with `poly` or `sigmoid` kernels.

Overall, we had the best parameters combination for all four metrics (precision, recall, f1-score, and Jaccard) at **0.988** (or 98.8%) with +/-0.013 stdv, except for Jaccard index metric where the best score was 0.976 (or 97.6%) with +/-0.025 stdv.

The best parameters combination for all metrics was as follows:

- `{'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}`

However, it's worth noting that the following parameters combination also scored very close to the one stated above:

- `{'C': 10, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}`
- `{'C': 100, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}`
- `{'C': 100, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}`
- `{'C': 1000, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': True}`
- `{'C': 1000, 'gamma': 0.001, 'kernel': 'rbf', 'shrinking': False}`

We can see that the parameters `C` and `shrinking` had little to no impact on this model and dataset with a Radial Basis Function (`rbf`) kernel; most of the impact is caused by the parameter `gamma` .

Part 2: Classification using multinomial logistic regression

Generate a classification system for this problem using *multinomial Logistic Regression*. Refer to the `sklearn_grid_search_digits.ipynb` Jupyter notebook example we discussed in class to tune the parameters for your estimator before building a model.

Solution

recall that multinomial logistic regression is simply a binary logistic regression classifier for multiple classes instead of 2. This can be done by deviding the problem into subproblems, where each subproblem is classifying 2 classes (or 2 group of classes). This will allow us to apply binary logistic regression classifier, in a nested/recursive fashion, to solve a multi-class classification problem.

first, let's create an MLR classifier without grid search and see how it performs with minimal variables specified.

Here we are only specifying the `multi_class` to specifically account for the multiclass classification problem at hand (digits 0..9), and `max_iter` as the default value (i.e., 100) causes a convergence warning due to reaching the maximum iteration limit and stopping prematurely.

```

first, let's create an MLR classifier without grid search and see how it performs
with minimal variables specified. Here we are only specifying the multi_class to
specifically account for the multiclass classification problem at hand (digits 0..9),
and max_iter as the default value (i.e., 100) causes a convergence warning due to
reaching the maximum iteration limit and stopping prematurely.
'''

```

```

# create multinomial logistic regression model
classifier = LogisticRegression(multi_class='multinomial', max_iter=1e4)

# train model
classifier.fit(X_train, y_train)

# predict target values using our trained model
y_pred = classifier.predict(X_test)

# show classification report
print(classification_report(y_test, y_pred))

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.96	0.95	0.95	91
2	0.97	1.00	0.98	88
3	1.00	0.92	0.96	92
4	0.95	0.99	0.97	91
5	0.91	0.98	0.94	91
6	1.00	0.98	0.99	91
7	1.00	0.96	0.98	89
8	0.93	0.91	0.92	87
9	0.93	0.96	0.95	90
accuracy			0.96	899
macro avg	0.96	0.96	0.96	899
weighted avg	0.96	0.96	0.96	899

Now let's try to create the same type of classifier, but utilize grid search to tune the classifier's hyperparameters

```

In [ ]: '''
now let's try to create the same type of classifier, but utilize grid search to tune
the classifier's hyperparameters
'''

```



```

# grid search parameters range and search cases
# See all possible parameters and their types, possible combinations, and their impact:
# https://scikit-Learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
searchCases = [
    {"multi_class": ["multinomial"],
     "penalty": ["l2"],
     "solver": ["newton-cg", "lbfgs", "sag", "saga"],
     "random_state": [1, 7, 77],
     "fit_intercept": [True, False],
     "C": [1, 10, 1000],
     "tol": [1e-4, 1e1, 1e2],
     "max_iter": [1e5],
     "warm_start": [True, False]
    },
    {"multi_class": ["multinomial"],
     "penalty": ["none"],
     "solver": ["newton-cg", "lbfgs", "sag", "saga"],
     "random_state": [1, 7, 77],
     "fit_intercept": [True, False],
     "tol": [1e-4, 1e1, 1e2],
     "max_iter": [1e5],
     "warm_start": [True, False]
    }
]

```

These are the metrics we will use to rank each grid search result and find the best possible outcome

```

In [ ]: # these are the metrics we will use to rank each grid search result and
# find the best possible outcome
metrics = ["precision", "recall", "f1", "jaccard"]
allResultsAllMetrics = []

for metric in metrics:
    allResultsPerMetric = []
    print(f"Finding the best hyperparameters for our MLR classifier based on the best {metric} score")

    # create model, and use grid search to tune the hyperparameters
    classifier = GridSearchCV(LogisticRegression(), searchCases, scoring=f"{metric}_macro")

    # train model
    classifier.fit(X_train, y_train)

    # display best results found via grid search
    print(f"Best parameters set found on development set:\n{classifier.best_params_}\n\n")

```

```

# display best parameters combination
means = classifier.cv_results_["mean_test_score"]
stds = classifier.cv_results_["std_test_score"]
for mean, std, params in zip(means, stds, classifier.cv_results_["params"]):
    allResultsPerMetric.append("%0.3f (+/-%0.03f) for %r" % (mean, std * 2, params))

# display detailed report of best case
print("Detailed classification report:\n")
print("The model is trained on the full development set.\n",
      "The scores are computed on the full evaluation set.\n")
y_true, y_pred = y_test, classifier.predict(X_test)
print(classification_report(y_true, y_pred), '\n\n')

# plot confusion matrix
fig, ax = plt.subplots(2, 5, figsize=(10, 6))
for cfs_matrix, target, axes in zip(multilabel_confusion_matrix(y_true, y_pred),
                                   range(0,10),
                                   ax.flatten()):
    addSubPlotConfusionMatrix(cfs_matrix, axes, target)
fig.suptitle(f'Multiclass Confusion Matrix for metric {metric}', fontsize=18)
fig.tight_layout(rect=[0, 0.03, 1, 0.95])
plt.show()
plt.clf()
print()

# append all result to the array to print it at the end
allResultsAllMetrics.append(allResultsPerMetric)

```

Finding the best hyperparameters for our MLR classifier based on the best precision score

Best parameters set found on development set:

```
{'C': 1, 'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 7, 'solver': 'saga', 'tol': 0.0001, 'warm_start': True}
```

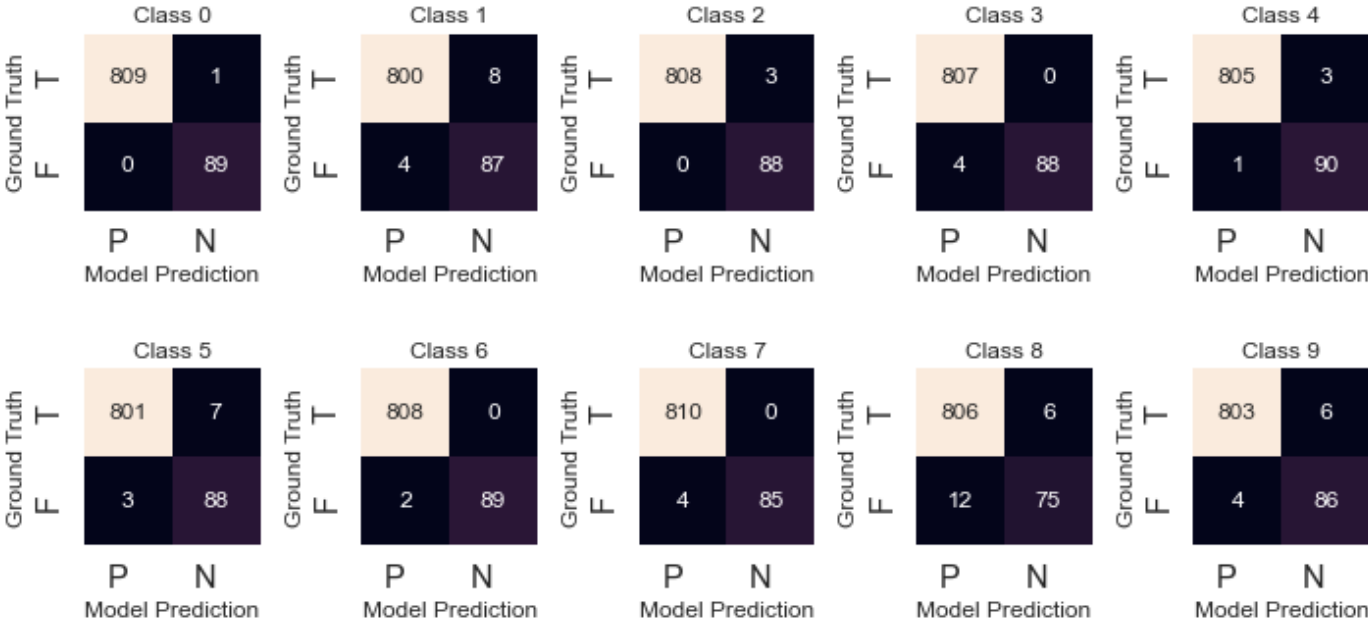
Detailed classification report:

The model is trained on the full development set.

The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	89
1	0.92	0.96	0.94	91
2	0.97	1.00	0.98	88
3	1.00	0.96	0.98	92
4	0.97	0.99	0.98	91
5	0.93	0.97	0.95	91
6	1.00	0.98	0.99	91
7	1.00	0.96	0.98	89
8	0.93	0.86	0.89	87
9	0.93	0.96	0.95	90
accuracy			0.96	899
macro avg	0.96	0.96	0.96	899
weighted avg	0.96	0.96	0.96	899

Multiclass Confusion Matrix for metric precision



Finding the best hyperparameters for our MLR classifier based on the best recall score
Best parameters set found on development set:
{'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': True}

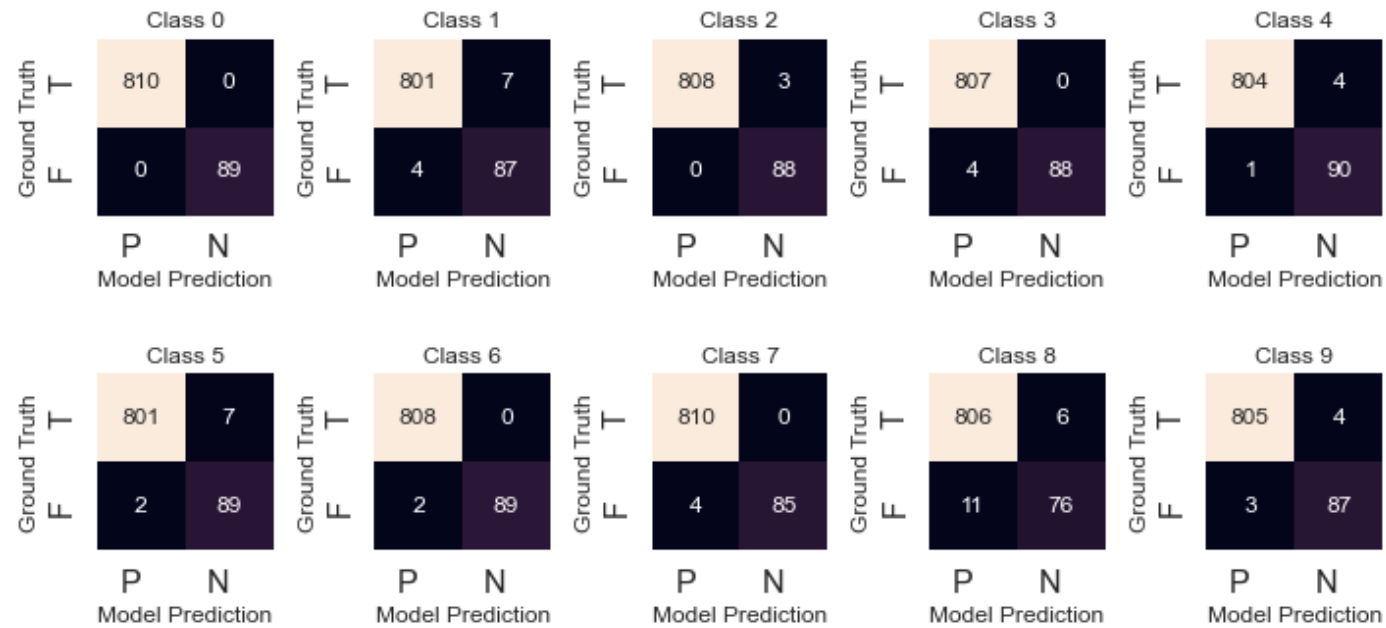
Detailed classification report:

The model is trained on the full development set.
The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.93	0.96	0.94	91
2	0.97	1.00	0.98	88
3	1.00	0.96	0.98	92
4	0.96	0.99	0.97	91
5	0.93	0.98	0.95	91
6	1.00	0.98	0.99	91
7	1.00	0.96	0.98	89
8	0.93	0.87	0.90	87
9	0.96	0.97	0.96	90
accuracy			0.97	899
macro avg	0.97	0.97	0.97	899
weighted avg	0.97	0.97	0.97	899

<Figure size 432x288 with 0 Axes>

Multiclass Confusion Matrix for metric recall



Finding the best hyperparameters for our MLR classifier based on the best f1 score

Best parameters set found on development set:

```
{'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': True}
```

Detailed classification report:

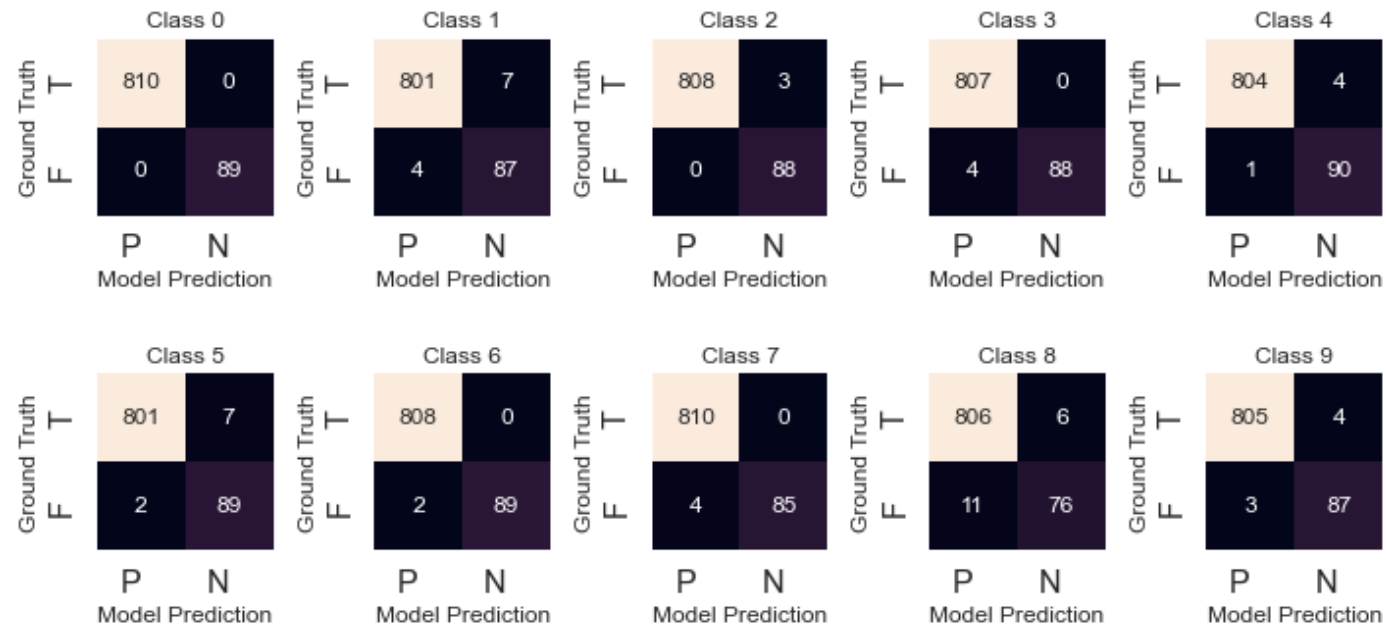
The model is trained on the full development set.

The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.93	0.96	0.94	91
2	0.97	1.00	0.98	88
3	1.00	0.96	0.98	92
4	0.96	0.99	0.97	91
5	0.93	0.98	0.95	91
6	1.00	0.98	0.99	91
7	1.00	0.96	0.98	89
8	0.93	0.87	0.90	87
9	0.96	0.97	0.96	90
accuracy			0.97	899
macro avg	0.97	0.97	0.97	899
weighted avg	0.97	0.97	0.97	899

<Figure size 432x288 with 0 Axes>

Multiclass Confusion Matrix for metric f1



Finding the best hyperparameters for our MLR classifier based on the best jaccard score

Best parameters set found on development set:

```
{'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': True}
```

Detailed classification report:

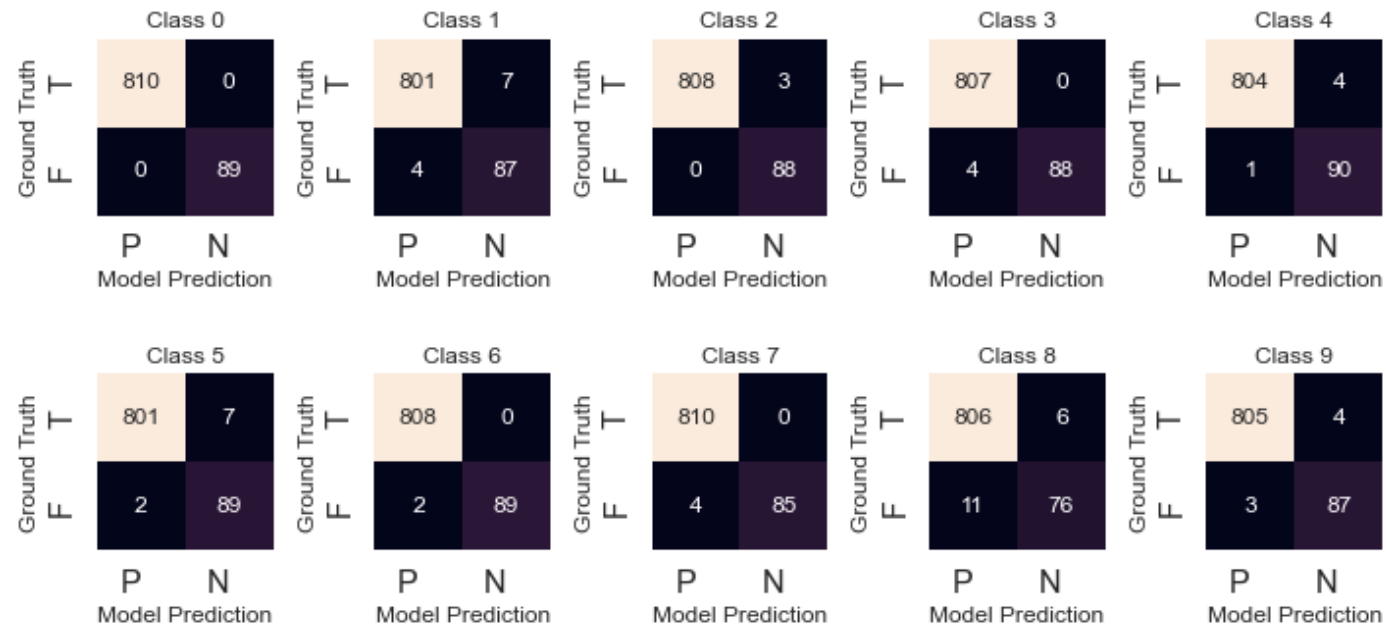
The model is trained on the full development set.

The scores are computed on the full evaluation set.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	89
1	0.93	0.96	0.94	91
2	0.97	1.00	0.98	88
3	1.00	0.96	0.98	92
4	0.96	0.99	0.97	91
5	0.93	0.98	0.95	91
6	1.00	0.98	0.99	91
7	1.00	0.96	0.98	89
8	0.93	0.87	0.90	87
9	0.96	0.97	0.96	90
accuracy			0.97	899
macro avg	0.97	0.97	0.97	899
weighted avg	0.97	0.97	0.97	899

<Figure size 432x288 with 0 Axes>

Multiclass Confusion Matrix for metric jaccard



<Figure size 432x288 with 0 Axes>

Printed all results for all possible results from our grid search for MLR classifier based on precision, recall, f1 score and jaccard

```
In [ ]: # print all results (full verbosity)
for metric in range(len(allResultsAllMetrics)):
    print(f"All possible results from our grid search for MLR classifier based on {metrics[metric]}")
    for x in allResultsAllMetrics[metric]:
        print('\t', x)
    print()
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

0.951 (+/-0.018) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 100.0, 'warm_start': True}
0.951 (+/-0.018) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 100.0, 'warm_start': False}
0.963 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 0.0001, 'warm_start': True}
0.963 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 0.0001, 'warm_start': False}
0.905 (+/-0.070) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 10.0, 'warm_start': True}
0.905 (+/-0.070) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 10.0, 'warm_start': False}
0.905 (+/-0.070) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 100.0, 'warm_start': True}
0.905 (+/-0.070) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 100.0, 'warm_start': False}
0.962 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': True}
0.962 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': False}
0.926 (+/-0.032) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': True}
0.926 (+/-0.032) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': False}
0.926 (+/-0.032) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': True}
0.926 (+/-0.032) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': False}

```

All possible results from our grid search for MLR classifier based on recall

```

0.958 (+/-0.021) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'newton-cg', 'tol': 0.0001, 'warm_start': True}
0.958 (+/-0.021) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'newton-cg', 'tol': 0.0001, 'warm_start': False}
0.961 (+/-0.031) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': True}
0.961 (+/-0.031) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': False}
0.950 (+/-0.010) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'newton-cg', 'tol': 100.0, 'warm_start': True}
0.950 (+/-0.010) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'newton-cg', 'tol': 100.0, 'warm_start': False}
0.958 (+/-0.021) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'lbfgs', 'tol': 0.0001, 'warm_start': True}
0.958 (+/-0.021) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

0.959 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'random_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': True}
0.959 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'random_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': False}
0.915 (+/-0.041) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'random_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': True}
0.915 (+/-0.041) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'random_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': False}
0.915 (+/-0.041) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'random_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': True}
0.915 (+/-0.041) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'random_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': False}

```

All possible results from our grid search for MLR classifier based on f1

```

0.958 (+/-0.022) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 0.0001, 'warm_start': True}
0.958 (+/-0.022) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 0.0001, 'warm_start': False}
0.961 (+/-0.031) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': True}
0.961 (+/-0.031) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': False}
0.950 (+/-0.010) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 100.0, 'warm_start': True}
0.950 (+/-0.010) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 100.0, 'warm_start': False}
0.958 (+/-0.022) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'lbfgs', 'tol': 0.0001, 'warm_start': True}
0.958 (+/-0.022) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'lbfgs', 'tol': 0.0001, 'warm_start': False}
0.955 (+/-0.041) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'lbfgs', 'tol': 10.0, 'warm_start': True}
0.955 (+/-0.041) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'lbfgs', 'tol': 10.0, 'warm_start': False}
0.946 (+/-0.022) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'lbfgs', 'tol': 100.0, 'warm_start': True}
0.946 (+/-0.022) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'lbfgs', 'tol': 100.0, 'warm_start': False}
0.959 (+/-0.011) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'sag', 'tol': 0.0001, 'warm_start': True}
0.959 (+/-0.011) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'sag', 'tol': 0.0001, 'warm_start': False}
0.888 (+/-0.088) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'sag', 'tol': 10.0, 'warm_start': True}
0.888 (+/-0.088) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

om_state': 77, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': False}
0.950 (+/-0.010) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'newton-cg', 'tol': 100.0, 'warm_start': True}
0.950 (+/-0.010) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'newton-cg', 'tol': 100.0, 'warm_start': False}
0.949 (+/-0.044) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 0.0001, 'warm_start': True}
0.949 (+/-0.044) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 0.0001, 'warm_start': False}
0.949 (+/-0.052) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 10.0, 'warm_start': True}
0.949 (+/-0.052) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 10.0, 'warm_start': False}
0.946 (+/-0.022) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 100.0, 'warm_start': True}
0.946 (+/-0.022) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'lbfgs', 'tol': 100.0, 'warm_start': False}
0.960 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 0.0001, 'warm_start': True}
0.960 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 0.0001, 'warm_start': False}
0.887 (+/-0.082) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 10.0, 'warm_start': True}
0.887 (+/-0.082) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 10.0, 'warm_start': False}
0.887 (+/-0.082) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 100.0, 'warm_start': True}
0.887 (+/-0.082) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 100.0, 'warm_start': False}
0.959 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': True}
0.959 (+/-0.013) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': False}
0.917 (+/-0.039) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': True}
0.917 (+/-0.039) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': False}
0.917 (+/-0.039) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': True}
0.917 (+/-0.039) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': False}

```

All possible results from our grid search for MLR classifier based on jaccard

```

0.921 (+/-0.039) for {'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2',
'random_state': 1, 'solver': 'newton-cg', 'tol': 0.0001, 'warm_start': True}

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```

om_state': 77, 'solver': 'lbfgs', 'tol': 100.0, 'warm_start': False}
0.925 (+/-0.024) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 0.0001, 'warm_start': True}
0.925 (+/-0.024) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 0.0001, 'warm_start': False}
0.806 (+/-0.131) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 10.0, 'warm_start': True}
0.806 (+/-0.131) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 10.0, 'warm_start': False}
0.806 (+/-0.131) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 100.0, 'warm_start': True}
0.806 (+/-0.131) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'sag', 'tol': 100.0, 'warm_start': False}
0.923 (+/-0.025) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': True}
0.923 (+/-0.025) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 0.0001, 'warm_start': False}
0.852 (+/-0.065) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': True}
0.852 (+/-0.065) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 10.0, 'warm_start': False}
0.852 (+/-0.065) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': True}
0.852 (+/-0.065) for {'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'none', 'rand
om_state': 77, 'solver': 'saga', 'tol': 100.0, 'warm_start': False}

```

Observation

We first ran a simple MLR classification model, then we attempted to run it again using grid-search. Our grid search cases were set based on the penalty parameter, as it's not possible to have non-default value for parameter `C` with a penalty value of `none`.

The best MLR model based on **percesion** score was **0.964 (96.4%)** with **+/-0.015** stdv for the following hyperparameters:

- `{'C': 1, 'fit_intercept': False, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 7, 'solver': 'saga', 'tol': 0.0001, 'warm_start': True}`

The best MLR model based on **recall** and **f1-score** was **0.961 (96.1%)** with **+/-0.031** stdv for the following hyperparameters:

- `{'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': True}`

The best MLR model based on **jaccard/similarity coefficient** score was **0.927 (92.7%)** with **+/-0.056** stdv for the following hyperparameters:

- `{'C': 1, 'fit_intercept': True, 'max_iter': 100000.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'random_state': 1, 'solver': 'newton-cg', 'tol': 10.0, 'warm_start': True}`

Lets build and validate the best model found from grid search

create multinomial logistic regression model with 5-folds cross validation and train the model

```
In [ ]: ''' build and validate the best model found from grid search '''

# create multinomial logistic regression model with 5-folds cross validation
classifier = LogisticRegressionCV(cv=5, Cs=1,
                                fit_intercept=False,
                                max_iter=100000.0,
                                multi_class='multinomial',
                                penalty='l2',
                                random_state=7,
                                solver='saga',
                                tol=0.0001)

# train model
classifier.fit(X, y)

# predict target values using our trained model
y_pred = classifier.predict(X)

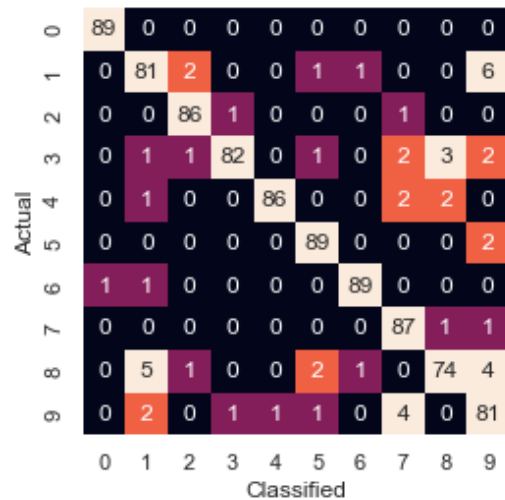
# show classification report
print(classification_report(y,y_pred))
```

	precision	recall	f1-score	support
0	0.98	0.99	0.99	178
1	0.87	0.90	0.88	182
2	0.97	0.96	0.96	177
3	0.99	0.91	0.95	183
4	0.98	0.96	0.97	181
5	0.96	0.95	0.96	182
6	0.98	0.97	0.97	181
7	0.92	0.99	0.95	179
8	0.90	0.84	0.87	174
9	0.85	0.91	0.88	180
accuracy			0.94	1797
macro avg	0.94	0.94	0.94	1797
weighted avg	0.94	0.94	0.94	1797

Lets plot the confusion matrix, Due to the high accuracy of this model, vmax parameter on the heatmap below set to 3 to make it easier to differentiate 1s from 0s and any value above 3 will appear as the same color.

```
In [ ]: expected = y_test
predicted = classifier.predict(X_test)

mat = confusion_matrix(expected, predicted)
sns.heatmap(mat, square=True, annot=True, fmt='d', cbar=False, vmax=3)
plt.xlabel('Classified')
plt.ylabel('Actual');
```



Part 3: Model comparison

Compare the two models. Which one did better and by how much?

Solution

The same dataset, and sub-datasets (post test and train) were used to train and test both models (SVM and MLR). This was to ensure that the comparison is normalized with the only variation being the classification models type.

Additionally, we utilized grid search, to ensure that we tuned each model's hyperparameters to give the best possible performance. Performance was measured based on four metrics: precision, recall, f1-score, and Jaccard.

Generally, the SVM classification model performed better than the MLR. More specifically, the best SVM model scored an accuracy of **99%**, whereas the best MLR model scored an accuracy of **96-97%** (based on grid search metric). In addition, MLR had more false positives than SVM in most classes (see generated classification matrix plots for each question).

Therefore, we can conclude that SVM did better than MLR by ~2% in accuracy.